The H1*−H2* Measure

Hyunkee Ahn*

ABSTRACT

In this paper, the H1*−H2* measure is introduced and exact procedures for obtaining the H1*−H2* value are fully specified. The H1*−H2* measure (a corrected difference in dB between the first and second harmonics) has been devised to provide an acoustic correlate of the phonation mode of a vowel following a consonant. With this measure, we can investigate the phonation mode of a vowel that is free from the F1 amplitude perturbation effect caused by the preceding consonant, which is especially salient at the voicing onset position of the vowel. For identical research purposes, on the other hand, the H1−H2 measure (the observed difference in dB between the first and second harmonic) has been employed in many previous studies. This paper compares these two measures by illustrating experimental results of exploring post−release phonation modes of vowels following the different manner classes of stop consonants in Korean i.e., the tense, lenis, and aspirated stops.

Keywords : H1*−H2* measure, H1−H2 measure, phonation, Korean stop sounds

1. Aims and Motivations of the Study

The main aim of this paper is to introduce a measure of H1*−H2*, a measurement of obtaining a corrected difference in dB between the first and second harmonics. This relatively unexplored measure will be specified in greater detail later. The theoretical values obtained by this measure are one way of determining a phonation mode of a vowel, which is free from the formant perturbation effects caused by the preceding consonant. Specifically, this measure can be employed, for example, to investigate the research question of whether the phonation mode of a vowel varies systematically according to the manner class of the preceding stop sound, especially at the voicing onset of the vowel. For the identical research question, the H1−H2 measure—i.e., the observed difference in dB between the first and second harmonic−−has been employed in previous studies (cf. Cho

---

* School of English Studies, Kyung Hee University.

et al., 2000, among others). However, the H1-H2 measure may not be entirely reliable if it is measured at the voicing onset of a vowel because of the salient F1 amplitude perturbation effects at that particular position. For the detailed comparison of these two measures, this paper will demonstrate the experimental results obtained by investigating phonation modes of vowels following the three phonemically different Korean stop consonants: the tense, lenis, and aspirated stops.

In actuality, it is in Stevens and Hanson(1995:160) that the H1*-H2* measure was introduced first. Their article, however, did not specify the exact procedures for obtaining the values of H1*-H2*. This is another motivation for me to launch this paper, since this measure needs to be fully specified anyhow for those who want to take advantage of this measure for their research.

2. Observed Measure of H1-H2

One way of determining a phonation type is by numeric measurements of the amplitude difference between the first and second harmonics (=Obs(H1-H2), henceforth). As described in Johnson (1997:127-30), the value of H1-H2 can play an important role as an index of the relative breathiness or creakiness of phonation. The general assumption is that the value of H1-H2 is larger in the breathy voice than that of the modal phonation, and the H1-H2 value of the modal phonation is larger than that of the creaky voice.

The difference in H1-H2 is mainly due to the difference in the shape of the glottal waveform. Specifically, the amplitude of the first harmonic in breathy phonation significantly dominates over the others given that this glottal waveform is, among the three phonations, most like a sine wave. Notice that the glottis is not completely closed even during the closing phase, causing the glottal waveform to be similar to a sine wave. The creaky phonation, on the other hand, does not show a difference in amplitude between the first few harmonics. Thus, the creaky waveform is least like a sine wave among the three phonations due to the closing phase of the glottal waveform, which falls off abruptly. The shape of the spectral envelope of the modal phonation is in-between.

If the spectral characteristics of the glottal waveform are directly reflected in the acoustic characteristics of a vowel, the value of H1-H2 can be used as an indication of determining the phonation type of that vowel. In this respect, the vowel in question should be a low vowel like [a], where the first formant (=F1), which is highest among the vowels, does not boost either the first or second harmonics appreciably.

3. Corrected Measure of H1*−H2*

However, the H1−H2 measure is not entirely reliable if it is used at the voicing onset of a vowel in a /CV/ context, where C is a stop sound, for example. This is because the first and second harmonics undergo a 'boost effect' due to the first formant during the transitional segment in the initial part of the vowel that follows a stop. Suppose, for example, that the first formant rises from 200 Hz at the voicing onset to a stable value of 800Hz. This formant transition would affect the amplitude levels of the first few harmonics at the voicing onset position. This amplification effect is clearly exemplified by Fant (1960:54−55), who has shown that the downward shift in F1 frequency with the rest of the formants being fixed results in an amplitude loss in the overall spectral envelope of the vowel. In this respect, the measure of Obs(H1−H2) is not a direct measure of source characteristics only. If the main concern of a study is to observe the phonation type realized at the voicing onset of the vowel following a stop, it should consider the fact that the laryngeal influence of the preceding stop is supposedly most salient at that particular position.

To correct for the first formant boosting effect on the first few harmonics at the voicing onset, Stevens and Hanson (1995) suggested a new method to normalize the amplitude difference between H1 and H2. Regarding this newly revised measure, Stevens and Hanson (1995:160) state that "Corrections were made for the amounts by which H1 and H2 are 'boosted' by the first formants, yielding the measure H1*−H2*."

The value of H1*−H2* is obtained by subtracting the predicted value of H1−H2 (=Exp(H1−H2), henceforth) from the observed value of H1−H2, as in formula (1) below.

(1) $H1^* - H2^* = Obs(H1-H2) - Exp(H1-H2)$

The Exp(H1−H2) is the predicted amplitude difference in dB's between the first and second harmonics.

According to the acoustic theory of speech production (Fant, 1960:49−60, 1972)), we can predict an expected value of Exp(H1−H2) if we know F_0 and the first few formant frequencies. This prediction is based on the assumption that the glottal wave is characteristic of modal phonation. Hence, the spectral tilt of the glottal source is fixed at −12 dB/*octave*. Notice that this spectral state of the glottal waveform is generally known to be a typical characteristic of so−called 'modal phonation'.. Since H1*−H2* compares observed and expected differences, it provides an indication of how the source spectrum deviates from the reference. In this respect, the value of H1*−H2* naturally represents a normalized amplitude difference between the first and second harmonics. For example, a zero value of H1*−H2* indicates that the sound wave observed at that particular time point has a glottal spectrum of modal phonation; specifically, the spectral tilt of the waveform falls off at a rate of -12 dB/*octave*.

3.1 Prediction of the 'Spectral Envelope' of Vowels

A classical statement of the 'Source-Filter' theory of speech production was presented in Fant (1960). According to this theory, all human speech sounds can be analyzed as the consequence of the generation of one or more sources of sound, and the filtering of these sources by the vocal tract (Stevens, 1999:56). The source-filter theory makes it possible to derive the 'spectral envelope' of vowels on the basis of the frequency values of the first few formants (Fant, 1960:48-60). Specifically, the spectral envelope of an arbitrary vowel is predictable given the formant frequencies. Other determinants of the spectral envelope, such as the bandwidth of each formant, the source characteristics, sound radiation, and higher formants (usually formants greater than F4), can be calculated from individual formulas. Notice, however, that his theory is based on a fixed spectral tilt of modal phonation at a rate of -12dB/*octave*.

First of all, the general source and filter formulation for voiced sounds is shown in (2):

(2) Sound = Source * Vocal Filter * Radiation Filter

$$|P(f)| = |U(f)|\,|H(f)|\,|R(f)|$$

where, brackets indicate absolute values, (f) indicates frequency.

Here, $|P(f)|$ symbolizes the output which is characterized by its amplitude-versus-frequency (spectral) properties. This function is analyzed as a product of three different frequency dependent functions: $|U(f)|$ the source spectrum; $|H(f)$ the vocal transfer function, and $|R(f)|$ the frequency characteristics of radiation (the conversion from volume velocity through the lips to pressure in the sound field). The combined functions of $|H(f)|$ and $|R(f)|$ thus constitute the complete filter function, $T(f)$. Therefore, we can rewrite the formula in (2) as $|P(f)| = |U(f)|\,|T(f)|$. This formula directly denotes the 'source-filter' theory of speech production in a mathematical manner. According to the standard variant of the theory, the two terms $|U(f)|$ and $|R(f)|$ can be predicted and calculated on the assumption that the source is idealized as the source associated with modal phonation. Its spectrum envelope will then fall off at a rate of -12 dB/*octave* at frequencies above 100 Hz. It has been shown that the spectral tilt of radiation is characterized by a +6 dB/*octave* rise, such that the combined source and radiation characteristics constitute a spectrum that falls off at the rate of -6dB/*octave* (= -12dB/*octave* +6 dB/*octave*). This relation is written in (3):

(3) Source and Radiation

$$|U(f)|\,|R(f)| = P_k \frac{f/100}{1+(f/100)^2}$$

where, P_k is a constant determining the particular sound pressure level. Note that the constant P_k is assumed to be equal to 1 in normal situation.

Next, the vocal tract transfer function $| H(f) |$ can be broken down into two basic components. One is the combination of the first four formants. The other accounts for the formants higher than F4. The analysis can be stated as follows:

(4) Vocal Tract Transfer Function

$$| H(f) | = k_{r4}(f) \; | H_1(f) | \, | H_2(f) | \, | H_3(f) | \, | H_4(f) |$$

In (4), $| H_1(f) |$ is the contribution from the first formant, $| H_2(f) |$ from the second formant, and so on. The factor $k_{r4}(f)$ contains the remaining frequency characteristics higher than the fourth formant and can be calculated based on the following formula:

(5) Contribution of formants higher than F4

$$20 \log_{10} k_{r4} = 0.54\, x^2 + 0.00143\, x^4 \text{ (dB)}$$

where, $x = f/f_1$ and $f_1 = 500$

Notice that the constant number, 500, is derived from a formula $f_1 = c/4l_{tot}$ where c denotes the speed of sound (=350m/sec) and l_{tot}, the total length of the vocal tract (=17.5 cm). The value of the contribution of each formant is calculated according to the following formula in (6):

(6) Contributions of F1 to F4

$$H_n(f) = \frac{F_n^2 + (B_n/2)^2}{\sqrt{(f - F_n)^2 + (B_n/2)^2}\sqrt{(f + F_n)^2 + (B_n/2)^2}}$$

where B_n denotes the bandwidth of F_n.

In Fant (1960:54), the value of each B_n was fixed at 100Hz, which is a simplification. To correct these, he later suggested formulas for calculation of B_1 – B_4 (Fant, 1972:47). The formulas are as seen in (7). Unlike B_1 – B_3, the value of B_4 is fixed at 300 Hz as seen in (7d).

(7) a. $B_1 = 15(500/F_1)^2 + 20(F_1/500)^{0.5} + 5(F_1/500)^2$ (Hz)

b. $B_2 = 22 + 16(F_1/500)^2 + 12000/(F_3 - F_2)$ (Hz)

c. $B_3 = 25(F_1/500)^2 + 4(F_2/500)^2 + 10F_3/(F_4 - F_3)$ (Hz)

d. $B_4 = 300$ (Hz)

The following formula in (8) is a summation of spectral levels of a particular vowel. It accounts for all the factors described so far. It describes a vowel as a spectral sum of four formants, a higher pole correction (a corrected value for formants higher than F4), and a combined source and radiation spectrum.

(8) Summation of spectral levels in dB

$$20 \log_{10} |P(f)| = 20 \log_{10} k_{r4}(f) + 20 \log_{10} |U(f)\, R(f)| + 20 \log_{10} |H_1(f)| + 20 \log_{10} |H_2(f)| + 20 \log_{10} |H_3(f)| + 20 \log_{10} |H_4(f)|$$

Specifically, the sub-formula $20 \log_{10} k_{r4}(f)$ is calculated according to (5), and $20 \log_{10} |U(f)\, R(f)|$ is calculated using (3). Formulas of (6) and (7) in combination help derive the value of $20 \log_{10} |H_n(f)|$.

3.2 Expected Value of (H1-H2)

The preceding remarks give us some insight into the structure of a vowel spectrum. With the aid of the several formulas introduced above, the value of Exp(H1-H2) is easily calculated, given the values of F_0 and the F-pattern. If we substitute the value of F_0 for the variable f in (8), for example, its amplitude value (=expected amplitude value of the first harmonic, H1) can be derived. A similar calculation is made for H2.

The H1*-H2* value is free from the variations of the F-patterns. Conversely, a value of Exp (H1-H2) varies depending on the F-pattern, so that the value naturally reflects the ‘boost-up’ effect. The value of H1*-H2* also refers to how much the phonation type of the target vowel is differentiated from the idealized modal phonation, which is assumed in the calculation of Exp(H1-H2). In principle, a negative value of H1*-H2* indicates that the glottal phonation in question is more like creaky voice, while a positive value of H1*-H2* means that the glottal phonation is closer to breathy voice. Overall, the Exp(H1-H2) measure serves as a fixed anchor, making the H1*-H2* measure meaningful and intelligible for acoustic studies of glottal phonation.

4. A Case Study: Post-Release Phonatory Processes in Korean

In this section, this paper investigates post-release phonation modes of vowels following Korean stop consonants, specifically in /_a/ sequences. Korean stops consist of the tense stops /p’, t’, k’/, the lenis stops /p, t, k/, and the aspirated stops /p^h, t^h, k^h/. The investigation makes significant use of the H1*-H2* measure during the initial portion of the vowel in order to provide an acoustic correlate of the post-release mode of phonation. By pursuing this research aim, we can naturally provide an example of comparing the two measures, the Obs(H1-H2) and H1*-H2* measures.

4.1 Experimental Methods

A total of 6 Korean male subjects participated in the recording. At the time of recordings, all subjects were graduate students attending the University of Texas at Austin (except for one of the subjects, who was a visiting scholar). None reported any medical problems influencing their language ability. The average age of the subjects was 36.5. They all speak standard Korean (Seoul dialect).

Speech samples were of CV structure with C being a stop consonant varying in place and manner and V being a fixed vowel [a]. Some of these items turned out to be real words, others nonsense words. For the data, the words in (9) were used, embedded in the carrier sentence in (10):

(9) a. tense series: /p'a/, /t'a/, /k'a/
b. lenis series: /pa/, /ta/, /ka/
c. aspirated series: /pʰa/, /tʰa/, /kʰa/

(10) Carrier sentence
sentence: ikesi _____ ita (Yale Romanization).
gloss: this + thing + nominative marker _____ +be (declarative form)
meaning: This is _______.

The subjects were required to repeat each of the items in (9) in succession until 5 clear tokens of each sample were obtained. Eventually, a total of 45 tokens was obtained from each subject (i.e., 3 manner categories * 3 places * 5 repetitions = 45 tokens). Subjects were recorded in a soundproofed room in the phonetics laboratory of the University of Texas at Austin. They were asked to speak the samples at normal speed and as naturally as possible in front of the microphone(Electro-Voice® 671A, Dynamic Cardioid, Electro-Voice, Inc.). The microphone was connected to a Power Mac computer (7100/80) via a stereo mixing console (Realisitc®, Model No. 32-1200 B). The recording for each subject took approximately 30 to 45 minutes.

Since it was important to keep the amplitude level of each token constant, a method of on-line digitization was adopted. The digitization was made at a sampling rate of 22,050 Hz with the aid of 'Sound Scope 1.43 f (Macintosh software program from GW Instruments, Inc.)'. Those signals clipped either at the top or bottom were discarded. The amplitude level of each token was easily maintained within the range of +/- 10 volts. In addition, when the subject found the pronunciation of the token unnatural, that token was also discarded. Some subjects produced the speech sample more than five times in a row until five clear signals were obtained. The digitized tokens were analyzed using Sound Scope to obtain the following raw data in (11).

(11) a. Amplitude levels of harmonic 1 and 2
b. F_0
C. Frequency values of formant 1 through formant 4

To obtain the values of the various measures in (11 a, b, c), a digital signal program of 'Fast Fourier Transform Routine' (=FFT, henceforth) included in Sound Scope was used with the following parameters in (12):

(12)	a. FFT points:	1024
	b. Bandwidth of Filter:	59 Hz (25 ms window)
	c. 6 dB pre-emphasis:	Off

The phonation mode (via H1*-H2*) and F_0 pattern at voicing onset position were most highly affected by the laryngeal settings of the preceding stop. In actuality, however, we obtained the relevant values at 13 ms away from the voicing onset of the vowel. The reason for measuring at +13ms point and not right on the first glottal pulse of a vowel (i.e., zero time point of a vowel) was that since the relevant FFT points centered around the marker on the waveform in this particular software program, and since the window frame is fixed at 25ms, the +13ms point (i.e., around half of 25 ms window) could be the minimum distance used to identify a phonation mode of a pure vowel at its earliest position measurable. If the marker of that program was on the voicing onset position, it would include a mixture of the sound of aspiration and the vowel.

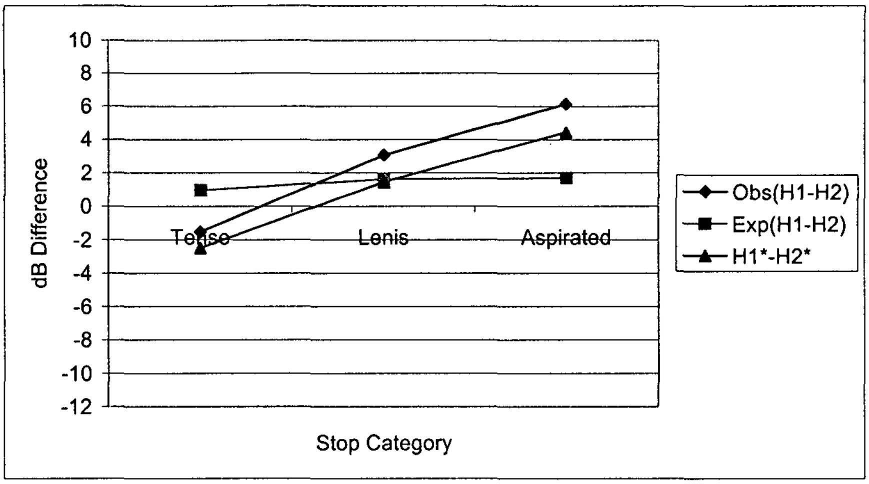
4.2 Results and Discussions

The numerical values in table 1 represents mean values of the Obs(H1-H2), Exp(H1-H2), and H1*-H2* measures for the three stop categories obtained at the +13 ms interval following the vowel onset. A line graph in figure 1 represents the the mean values. For the detailed graphical and statistical results, see Ahn (1999).

Table 1. Mean values of the Obs(H1-H2), Exp(H1-H2), and H1*-H2* measures at the +13ms time point

Manner Classes	Obs(H1-H2)	Exp(H1-H2)	H1*-H2*
Tense	-1.55	0.95	-2.50
Lenis	3.06	1.61	1.45
Aspirated	6.13	1.69	4.44

Figure 1. Line graph with average data points for Obs(H1-H2), Exp(H1-H2), and H1*-H2* for the three stop categories. The observations were obtained at the +13 ms time point and were averaged over all 6 Korean subjects.



The F1 amplitude perturbation effects are reflected in the Exp(H1−H2) values. In other words, the Exp(H1−H2) pattern reflects the predicted effects on the first two harmonics of the first formant. Not surprisingly, the post−lenis and post−aspirated classes show somewhat higher values in this measure than the post−tense class. This is mainly because of the formant transition taking place during the long aspiration interval. The long duration of aspiration, which includes much part of formant transition, places the formant measurements later as is usual for stops with long VOT's. Hence there is no significant influence on the value of Exp(H1−H2) in post−aspirated and post−lenis classes—that is, the measurement at the +13ms time point is somewhat free from the F1 amplitude perturbation effect. Due to their short VOT, on the other hand, the post−tense class shows Exp(H1−H2) values as smaller than those of the other two classes. As for the post−tense class, the F1 transition is salient in the initial portion of the vowel, causing a F1 amplitude perturbation effect on the first and second harmonics (especially on the second harmonic). That is, the amplitudes of the first and second harmonics are boosted by the F1 transition for this class. This is why the value of Exp(H1−H2) of the post−tense class is smaller at the +13ms time point than those for the other two classes. For reference, table 2 shows the average values of VOT for the three classes.

Table 2. Descriptive statistics for the VOT measure based on the Korean data

Manner Class	Mean Value (ms)	Standard Deviation (ms)
Unaspirated	13.46	7.81
Voiced	59.40	21.23
Aspirated	128.52	27.99

The values obtained by the H1*−H2* measure have some more interesting phonetic advantages over those values obtained by the Obs(H1−H2) measure. First, the H1*−H2* value, as fully discussed before, is free from the variations of the F-patterns, while the Obs(H1-H2) value is not. Second, the value of H1*-H2* reflects deviations of the phonation mode of the target vowel from the idealized modal phonation, while the value of Obs(H1-H2) does not. For example, the average H1*-H2* value, -2.50, in the post-tense class presents a considerable departure from the zero value of H1*-H2*, representing creaky-like phonation. On the other hand, the post-lenis and post-aspirated classes show somewhat breathy-like phonation, as revealed by the average values of 1.45 and 4.44, respectively. Thirdly, the H1*-H2* values at the +13 ms time point will indirectly allow us insight into the laryngeal settings made in the post-release circumstances without the use of fiberoptic measures of glottal width. The different articulatory gestures characterizing the three stop series in Korean have been described in previous studies (Kagaya, 1974, among others) as follows: constricted glottis for tense stops, spread glottis for aspirated stops, and/or moderate glottis for lenis stops. Therefore it is not totally unreasonable to assume that these physiological characteristics are reflected most saliently in the H1*-H2* values at this particular time point. However, since the Obs(H1-H2) values are not totally free from the F1 amplitude perturbation effects, it is suspicious to assume that the values directly reflect these physiological characteristics.

In spite of these afore-mentioned advantages of the H1*-H2* measure over the Obs(H1-H2) measure, there are some limitations on interpreting the theoretical H1*-H2* data, which result from the calculation of the Exp(H1-H2) data. Specifically, the theoretical Exp(H1-H2) values are calculated on the assumption that all male speakers have an identical vocal tract length of 17.5 cm, and that the sound speed is 350 m/s. Another potential weakness concerning the Exp(H1-H2) value comes from the assumption that the formant values can be derived by the frequency values of the relevant harmonics obtained by spectrum readings.

REFERENCES

Ahn, Hyunkee. 1999. *Post-Release Phonatory Processes in English and Korean: Acoustic Correlates and Implications for Korean Phonology.* Ph.D. dissertation. The University of Texas at Austin.

Cho, T., S-A., Jun and P. Ladefoged. 2000. "An Acoustic and Aerodynamic Study of Consoants in Cheju." Speech Sciences, 7:1, 109-141.

Fant, G. 1960. *Acoustic Theory of Speech Production.* Mouton: The Hague.

______1972. "Vocal Tract Wall Effects, Losses, and Resonance Bandwidths." *STL-QPSR,* 2-3, 28-52.

Johnson, K. 1997. *Acoustic and Auditory Phonetics.* Cambridge: Blackwell.

Kagaya, R. 1974. "A Fiberscopic and Acoustic Study of the Korean Stops, Affricates and

Fricatives." *Journal of Phonetics*, 2, 161–180.

Stevens, K. N. and H. M. Hanson. 1995. "Classification of Glottal Vibration from Acoustic Measurements." in O. Fujimura & M. Hirano, eds., *Vocal Fold Physiology: Voice Quality Control*, 147–170. San Diego: Singular.

Stevens, K. N. 1999. *Acoustic Phonetics*, Cambridge: The MIT Press.

▲ Hyunkee Ahn
School of English Studies, Kyung Hee University
1 Hoegi–dong, Dongdaemun–gu, Seoul 130–701, Korea
Tel : (02) 961–0211 (O), Fax : (02) 961–9317
H/P: 018–244–8983
e–mail: hkahn@nms.kyunghee.ac.kr